Demonstrating Correspondence Between Decision-Support Models and Dynamics of Real-World Environmental Systems

Ray Huffaker[1], Rafael Muñoz-Carpena[1,*], Miguel Campo-Bescos[2], Jane Southworth[3]

[1] Department of Agricultural and Biological Engineering, University of Florida, USA, 281 Frazier Rogers Hall, Gainesville, FL 32611-0570, USA,

[2] Projects and Rural Engineering Department, Public University of Navarre, Ed. Los Olivos, Pamplona, Spain

[3]Geography Department, University of Florida, Gainesville, Florida, USA

[*]Corresponding author: carpena@ufl.edu, (352)392-1864 Ext. 287 (phone), (352)392-4092 (fax).

Highlights

- A pre-modeling Nonlinear Time Series (NLTS) framework presented
- Framework reconstructs real-world system dynamics from observed data
- Reconstructed dynamics inform model specification and auditing

Abstract

There are increasing calls to audit decision-support models used for environmental policy to ensure that they correspond with the reality facing policy makers. Modelers can establish correspondence by providing empirical evidence of real-world behavior that their models skillfully simulate. Since real-world behavior—especially in environmental systems—is often complex, credibly modeling underlying dynamics is essential. We present a pre-modeling diagnostic framework based on *Nonlinear Time Series* (*NLTS*) methods for reconstructing real-world environmental dynamics from observed data. The framework is illustrated with a case study of saltwater intrusion into coastal wetlands in Everglades National Park, Florida, USA. We propose that environmental modelers test for systematic dynamic behavior in observed data before resorting to conventional stochastic exploratory approaches unable to detect this valuable information. Reconstructed data dynamics can be used, along with other expert information, as a rigorous benchmark to guide specification and testing of environmental decision-support models corresponding with real-world behavior.

Keywords: Model evaluation, Nonlinear dynamics, Phase space reconstruction, Extreme Value Statistics

1. Introduction

Real-world environmental systems are complex, ever-changing, and beyond anyone's capacity to model closely. Despite this, policymakers rely on decision-support models to regulate real-world environmental systems. Faulty model representations lead to ineffective and wasteful policies [1]. Consequently, there are increasing calls to review decision-support models used for public policy [1, 2]. For example, the European Commission's Joint Research Centre formally audits models used in impact assessments of EU initiatives, legislation, and policy [3].

If modelers are required to demonstrate correspondence of their models with real-world behavior—as recommended by Oreskes et al. (1994) [2] and Saltelli and Funtowitz (2014) [1]—what is a reasonable burden of proof? The literature is clear that modelers cannot be reasonably required to verify their models as accurate representations of reality since verification is a logical impossibility. Moreover, demonstrating a 'good fit' between model output and observed data does not constitute validation since the possibility remains that other models with very different structures and representations of reality can be parameterized to provide good fits [2, 4, 5]. Because real-world environmental systems are in a constant state of flux, we propose that a reasonable burden of proof would require auditing modelers to present persuasive empirical evidence of real-world dynamic behavior that their models skillfully simulate.

What constitutes persuasive evidence of real-world dynamic behavior? Most would agree that observed data provide an essential portal to real-world environmental systems to which there is only limited access. Evidence of real-world dynamic behavior must be drawn somehow from data that often exhibit a highly volatile, irregular, and random appearance.

Past work in the environmental and resource management literature presumes that irregular data reflect two major sources of uncertainty: (1) 'Real' uncertainty due to the inherent

randomness and natural variation of real-world biophysical processes, and (2) 'Perceived' uncertainty due to decision-maker's limited perceptions of reality [6-9]. Typically, stochastic decision-support models are believed to be required to capture this management uncertainty [6], while deterministic methods are believed to ignore or assume away this uncertainty, and thus incapable of representing risk-responsive management behavior [10].

Contrary to this common presumption, principles of randomness—set out in the philosophy of science literature—reveal that random-appearing data are not evidence that stochastic decision-support models are required to model uncertainty [11, 12]. Stochastic modeling is required only when real-world processes are 'physically random' (indeterministic). Since indeterministic processes are irreducibly probabilistic, causal relationships do not exist to support deterministic formulations. Alternatively, deterministic modeling is feasible when real-world processes are physically nonrandom (deterministic). Deterministic processes are governed by laws such that, given them, nothing else can happen. Importantly, apparent 'mathematically random' process output can be generated by both indeterministic and deterministic processes. Indeed, breakthroughs in nonlinear dynamics demonstrate that irregular and apparently-random dynamic behavior can emerge endogenously from deterministic, low-dimensional and nonlinear interactions among system variables [13, 14]. Equally importantly, one cannot work backward from observing mathematically random output to prove whether it was generated by an indeterministic or deterministic process.

In sum, random-appearing environmental data are not evidence of indeterministic biophysical processes requiring stochastic modeling. The interesting possibility remains that deterministic decision-support models might generate observed uncertainty endogenously. In this paper, we propose a pre-modeling diagnostic framework to empirically test observed data for

this possibility. The framework is based on *Nonlinear Time Series* (*NLTS*) methods—developed in mathematical physics—that reconstruct system dynamics from time-series data on a single variable [15]. The *NLTS* framework extends Larsen *et al.* (2014) who proposed an exploratory approach in which modelers experiment with simple nonlinear deterministic structures to find those capable of generating complex observed environmental dynamics [16]. We address the open question of how to diagnose whether real-world environmental dynamics are nonlinear and deterministic in the first place. *NLTS* data diagnostics provide mathematically and statistically rigorous evidence of real-world dynamic behavior, and can guide specification and testing of deterministic environmental decision-support models corresponding with diagnosed behavior.

We present an intuitive description of each component of the proposed *NLTS* data diagnostics framework, and illustrate it with a case study of saltwater intrusion into coastal wetlands in Everglades National Park, Florida, USA.

2. Framework for *NTLS* Pre-Modeling Data Diagnostics and Data-Informed Modeling

The proposed *NLTS* diagnostic framework is summarized in Figure 1. In a nutshell, we first apply *Singular Spectrum Analysis* (*SSA*) [17-19]—a signal processing technique—to separate an observed time series into signal (structured variation) and noise (unstructured variation). *Singular Spectrum Analysis* (*SSA*) is a data-adaptive signal processing approach that can accommodate highly anharmonic (potentially non-sinusoidal) oscillations in irregular signals [17-20]. When the time series is converted to anomalies from the mean, and the *Toeplitz* method of *SSA* is applied, signal strength is measured as the fraction of variation explained in the observed time series from its mean [18, 19]. We test a sufficiently strong signal for low-dimensional nonlinear dynamic structure with *Phase Space Reconstruction* [15, 21] and *Surrogate Data* methods [22]. We propose a novel application of *Extreme Value Statistics* [23]

to model noise separated from an observed time series probabilistically. This produces an informative diagnostic estimating the time expected before a particular noise level is realized.

Data testing positive for low-dimensional nonlinear dynamic structure open the door to a number of nonlinear data diagnostics that are intended to inform—not replace—subsequent mechanistic modeling. If multiple time series are available, *Convergent Cross Mapping* [24] empirically tests for causal networks, and *Phenomenological Modeling* [25] fits a system of ordinary differential equations with polynomial interaction terms in the causal network variables that reproduces dynamics empirically reconstructed from the data (*Data-driven modeling*). *State-Space Forecasting* [21] can make out-of-sample forecasts of the structured variation (signal) separated from the observed data. Alternatively, if an observed time series provides a weak signal, or a strong signal tests negative for low-dimensional dynamic structure, conventional linear stochastic methods provide a viable option. In either case, *NLTS* diagnostics replace presumption with empirical evidence to make a more compelling case that the modeling approach selected is well specified to represent the dynamics of real-world environmental systems facing decision-makers.

2.1 Signal/Noise Separation

The ability of *NLTS* to provide useful empirical evidence of real-world dynamics is restricted by the information contained in observed data. Available data are typically noisy due to, e.g., measurement, recording and processing errors. Noisy data increase the difficulty of detecting oscillatory patterns and decrease the resolution of reconstructed dynamics [26]. Signal processing isolates the structured variation in data and measures the extent to which it accounts for variation in the data.

In addition, available data sets are often short. This presents a challenge because, similar to

other time series methods, *NLTS* demands stationary data. Stationarity requires that the "duration of the measurement is long compared to the time scales of the systems." ([15], p. 33) This means that historic records must be of sufficient length to provide an adequate sampling of important oscillatory patterns occurring at lower frequencies. Otherwise, some lower-frequency behavior is grouped into linear or nonlinear trends. Conventional *NLTS* practice applies signal processing methods that detrend data to isolate detected oscillations [27]. Trend can be re-introduced in model design exogenously as a time-dependent parameter, or endogenously if low-frequency processes undetected by *NLTS* can be identified from other reliable information. Another option is to split the time series into shorter stationary segments. However, the risk is that each segment will be too short to adequately detect real-world dynamics.

2.2 Phase Space Reconstruction

A strong (detrended) signal is tested for low-dimensional nonlinear dynamic structure with phase-space methods [21]. Each point in phase space records the level of system variables in a given time period (the 'state' of the system). A unique trajectory passing through each point shows the co-evolution of system variables satisfying transition rules defined by a system of first-order differential equations. In *dissipative* dynamic systems, variables co-evolve from given initial conditions toward a system *attractor* bounded within a submanifold of phase space [21]. Attractors are geometric structures with "noticeable regularity" ([28], p. 55), and are *invariant* meaning that any trajectory starting on an attractor never leaves [29]. Examples include stable fixed points, stable limit cycles, and chaotic attractors upon which solution trajectories oscillate irregularly [29, 30]. Chaotic attractors exhibit *sensitivity dependence on initial conditions*. Points on close neighboring trajectories of an attractor exponentially diverge toward very different future points [29]. Accurate long-term predictions of chaotic systems are

unachievable. Despite exponential divergence, chaotic trajectories converge toward a bounded, spatially-organized, and low-dimensional 'strange attractor' upon which they orbit irregularly. Unlike a point (no dimension), a line (one dimension), or a surface (two dimensions), chaotic attractors have a non-integer *fractal* dimension [15, 31, 32].

Takens (1980) [33] proved that a 'shadow' version of an attractor can be reconstructed from a single system variable without knowing the system equations (*Phase Space Reconstruction*). For example, assume that a real-world dynamic system is comprised of three variables—*x(t)*, *y(t)*, and *z(t)*—and is characterized by a low-dimensional attractor, but that only a historical record of *x(t)* is observed. The *time-delay embedding* method replaces the unobservable original coordinate system (*x*, *y*, *z*) with three delayed copies of the single observed variable: *x(t)*, *x(t-d)*, and *x(t-2d)* with delay *d* [31]. The number of delay-coordinate vectors required to define the *shadow* phase space is termed the *embedding dimension* (*M*), which estimates the minimum system dimensionality required to contain a reconstructed attractor. The shadow attractor reconstructed from delayed copies of *x(t)* preserves essential mathematical properties of the original because it is a 1-1 mapping of dynamics from the original to the delay coordinates. The underlying intuition is that any single variable in an integrative system encodes the history of its interactions with other system variables [34].

The sufficient condition guaranteeing the Takens result is: $M \geq (2N+1)$, where *N* is the dimension of the submanifold bounding the original attractor [33]. Since this is a sufficient condition, fewer delay-coordinate vectors may be necessary to reconstruct a shadow attractor [35]. Moreover, since *N* is unobserved, the sufficient condition cannot be used to directly determine embedding dimension *M* in practice [32]. Instead, statistical methods are used to select an embedding delay (*d*) and embedding dimension (*M*). The embedding delay *d* is

conventionally selected as the first minimum of the 'mutual information function'—a probabilistic measure of the extent to which *x(t)* is related to *x(t-d)* at a given *d*. This selection is thought to introduce needed statistical independence between successive delayed values of *x(t)*. If the delay is too short, successive delayed values contain redundant mutual information, and the reconstructed attractor is restricted to rest on a diagonal in the embedding space. If the delay is too long, successive delayed values contain insufficient mutual information to reconstruct the attractor [32].

The embedding dimension *M* is conventionally chosen with the 'false nearest neighbors' method. This method measures the percentage of points that are close neighbors in a given dimension that remain close in the next highest dimension. Points growing apart in the next highest dimension are 'false' neighbors. The embedding dimension *M* is that for which the percentage of false neighbors falls to zero for a given tolerance level. A reconstructed attractor embedded into fewer dimensions would have insufficient space to fully express itself. An attractor embedded into more dimensions would begin to go out of focus [32].

There are important caveats to *Phase Space Reconstruction* when operating with small and noisy data sets [20]. One is limited to reconstructing a 'sampling' or 'skeleton' of a real-world attractor [18]. Moreover, one may fail to reconstruct even the skeleton of a shadow attractor for several reasons including: 1) system dynamics are not governed by a low-dimensional attractor; 2) noisy or limited data prevent detection of an existing attractor; or 3) observed data do not lie on an attractor [32].

2.3 Surrogate Data Testing

Reconstruction of a shadow attractor from a strong signal provides preliminary evidence of real-world low-dimensional nonlinear deterministic dynamics. The possibility remains that

noticeable regularity in an attractor is the figment of a mimicking linear stochastic process [22]. *Surrogate Data* methods statistically test the null hypothesis of mimicking linear stochastic dynamics [22, 35-37]. Rejecting the null hypothesis provides corroborative empirical evidence that visualized structure in an empirically-reconstructed attractor may be due to low-dimensional and nonlinear deterministic dynamics.

First, a set of surrogate data vectors is randomly generated from the (detrended) signal to destroy intertemporal patterns while preserving various statistical properties. The *Amplitude Adjusted Fourier Transform* (*AAFT*) algorithm—the most prevalent in application—calculates surrogate data as static monotonic nonlinear transformations of linearly filtered noise designed to preserve both the signal's probability distribution and power spectrum [22, 37, 38]. The *AAFT* algorithm has been criticized because the assumption of a static monotonic transform may not be met when working with real data [38]. As a result, *AAFT* surrogates may fail to reproduce linear correlations in the data [38], and may produce a biased (flatter) power spectrum [37].

Theiler et al. (1992) recommended that a more general algorithm be developed to test the null hypothesis that real-world dynamics are characterized by a noisy limit cycle [22]. The *Pseudo Phase Space* (*PPS*) algorithm tests the null hypothesis that nonrepeating cycling detected in an empirically-reconstructed attractor is most likely due to a randomly-shifting limit cycle characteristic of stochastic linear dynamics. It preserves periodic trends in the signal while destroying chaotic structures [35, 36].

Our strategy is to follow conventional practice in first testing the null hypothesis of linear stochastic dynamics with the *AAFT* algorithm, and then follow Kugiumtzis (2008) [39] to re-test with a more general algorithm such as the *PPS* algorithm.

Second, phase space is reconstructed for each surrogate vector, and selected *discriminating*

statistics measuring various attractor characteristics are estimated. Conventional statistics for detecting 'hallmarks' of deterministic structure are the *correlation dimension* (measuring the extent to which points on a reconstructed attractor are spatially organized), and the *Lyapunov Exponent* (measuring sensitivity to initial conditions and resultant spreading of state-space trajectories over time). Although these two discriminating statistics must be used cautiously in empirical work because of difficulties in computing reliable estimates from finite noisy records [15, 22], they can be used to reliably distinguish between deterministic and random structure in surrogate data tests [15].

Another conventional discriminating statistic is the skill with which a reconstructed attractor can forecast [31]. Following state-space forecasting methods, points on the reconstructed and surrogate attractors are split into forecasting and validation bases. Initially, the nearest neighboring points to the final point in the forecast base are computed, advanced one time period, and averaged to forecast the first point in the validation base. At each step, the forecasting base is augmented by a point in the validation base until all points in the validation base (excepting the final point) have been predicted [21, 31]. The first coordinate of the forecasted point on the reconstructed attractor is the forecasted signal, and the remaining coordinates are delayed forecasts. State-space forecasting also can be used to forecast the signal out-of-sample. The forecasting base then includes all points on the attractor, and at each step, the predicted point is added to the attractor to predict the next point out-of-sample [31].

Predictive skill can be calculated with a number of goodness-of-fit measures including *Nash-Sutcliffe Efficiency* [40-42]:

$$nse = 1 - \frac{\sum_{i=1}^{k}\left(x_k - x_{Fk}\right)^2}{\sum_{i=1}^{k}\left(x_k - \bar{x}\right)^2} = 1 - \left(\frac{RMSE}{s}\right)^2 \quad (1)$$

where k denotes periods in the validation base, x_k and x_{Fk} are the signal and its predicted value in period k, respectively, and $\bar{x}$ is the signal averaged over the validation base. *RMSE* is the root mean square error of the forecasted values against the observed data, and σ is the standard deviation of the observed data. A value $nse = 1$ represents a perfect fit and a value of $nse > 0.65$ is often proposed as a model quality threshold [41].

We test the null hypothesis of mimicking linear stochastic dynamics with nonparametric rank-order statistics [22, 37]. The number of surrogate data vectors generated for a two-tailed test is: $S = (2k/\alpha) - 1$, where α sets the probability of false rejection, $(1-\alpha)\times 100$ is the level of significance, and k is a positive integer such that larger values produce a more sensitive test. The null hypothesis is rejected if a discriminating statistic taken from the empirical-reconstructed attractor is among the k smallest or k largest values in the ensemble of statistics taken from the surrogate attractors. For a single-tailed test, $S = (k/\alpha) - 1$ surrogates are generated, and the null hypothesis is rejected if a discriminating statistic taken from the empirical-reconstructed attractor is among the k smallest (for a lower-tailed test) or k largest (for an upper-tailed test). We conduct a two-tailed test for correlation dimension and Lyapunov exponent, and a single-tailed test for predictive skill to reject the null hypothesis only if the empirically-reconstructed attractor predicts with more skill that its surrogate counterparts.

2.4 Causal Network Analysis

When data test positive for nonlinear dynamics, nonlinear feedback relationships encode information about variable *Y* into variable *X*, and this information is not lost by removing *Y* from

the system, i.e., nonlinear dynamics are intrinsically non-separable. Under these circumstances, *Granger Causality* and conventional correlation tests of causation provide ambiguous results [24]. Sugihara et al. (2013) recently developed *Convergent Cross Mapping* (*CCM*) to detect causality in non-separable, nonlinear and deterministic dynamic systems [24].

CCM uses the empirically-reconstructed attractors for two observed variables to test whether they belong to the same dynamic system, and thus causally interact. If X and Y belong to the same system, then an attractor reconstructed with either delayed X-coordinates (A_X) or delayed Y-coordinates (A_Y) both map 1-1 to the original attractor (A), and consequently map 1-1 to each other. To test for a 1-1 mapping between A_Y and A_X, the nearest neighbors to a given reference point on A_Y (for example) are identified and their time indices used to find contemporaneous points on A_X. These will also be neighboring points on A_X only if there is a 1-1 mapping with A_Y. This cross mapping can be used to predict states of X from the historic record for Y, and vice-versa. Cross-mapped predictions of X using A_Y that match well with corresponding observed values of X indicate that X and Y are causally related. A measure of the strength of causal interaction is that cross-mapped predictions converge toward the observed value as the portion of the historic record for Y used to reconstruct A_Y increases in length. The goodness-of-fit between the two can be measured with statistics including *Nash-Sutcliffe Efficiency* and the *Pearson Correlation Coefficient*. Stronger causal interaction is indicated as both statistics converge closer to one [24].

2.5 Phenomenological Modeling

Phenomenological modeling is used in applied nonlinear dynamics to reproduce empirically-detected dynamics with a system of first-order ordinary differential equations expressed as polynomial expansions of interacting observed variables [25, 43]:

$$\left.\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1(t), x_1(t), \ldots, x_n(t); \Omega_1) \\ \frac{dx_2}{dt} &= f_2(x_1(t), x_1(t), \ldots, x_n(t); \Omega_1) \\ &\ldots \\ \frac{dx_n}{dt} &= f_n(x_1(t), x_1(t), \ldots, x_n(t); \Omega_1) \end{aligned}\right\} (2)$$

where Ω_i $(i = 1,...n)$ are parameter vectors for each equation. We apply phenomenological modeling to analytically reproduce the empirically-reconstructed attractor. As explained below, this provides useful information for subsequent mechanistic modeling. The phenomenological model's dimensionality (n) is guided by the embedding dimension (M) estimated in *Phase Space Reconstruction*, and the selected variables by the causal network empirically detected with *CCM*.

Time derivatives are approximated with 4th order centered finite differences of the de-trended signals separated from each observed time series [44]. The order of each polynomial is selected so that the model faithfully reproduces the empirically-reconstructed attractor, and simulated variables reproduce the spectral properties of the observed signals. If the phenomenological model is three dimensional or lower, a simulated attractor can be constructed from a scatterplot of the simulated variables. If the model is greater than three dimensions, a simulated attractor can be reconstructed from one of the simulates with time-delay embedding. Model parameters Ω_i $(i = 1,...n)$ can be estimated with *Ordinary Least Squares* regression since the model is linear in parameters [43], or *Partial Least Squares* regression if needed to correct for collinearity among polynomial regressors [45].

2.6 Modeling Noise

Noise measures the discrepancy between an observed time series and separated signal, and consequently comprises events occurring outside detected behavioral patterns. Positive (negative) noise is the amount by which the observed time series exceeds (falls below) the signal.

The signal may include regularly occurring disasters such as floods, droughts or epidemics. However, disasters also may appear as noise if they occur irregularly, or with regularity that cannot be detected with limited available data. Since disaster mitigation is core to public policy protecting social welfare, it is essential to account for them whether they occur as signal or noise.

Extreme Value Statistics models extreme events that are statistically independent [23]—a condition likely met by unstructured noise residual to effective signal processing. We apply *Extreme Value Statistics* to calculate the likelihood of noise exceeding a selected threshold value. A conventional way to select a threshold is the level at which the *mean-residual-life* plot becomes linear with increasing thresholds [46]. Exceedances follow a Generalized Pareto (*GP*) distribution. A 'shape' parameter, ξ, determines whether the *GP* distribution conforms to a light-tailed exponential distribution $(\xi = 0)$, a heavy-tailed Pareto distribution $(\xi > 0)$, or a bounded beta distribution $(\xi < 0)$. Quantiles computed from the fitted distribution are inverted to produce 'return-level plots'—an informative noise diagnostic estimating the time periods ('return periods') expected before particular noise levels exceeding the selected threshold ('return levels') are realized.

2.7 Computational Packages

The following ***R*** packages are available to run the proposed *NLTS* methods: 'Rssa' (*Singular Spectrum Analysis*); 'tseriesChaos', 'nonlinearTseries', 'RTisean', and 'fractal' (*Phase Space Reconstruction* and *Surrogate Data Analysis*); 'multispatialCCM' (*Convergent Cross Mapping*); 'igraph' (network diagrams); 'lm' (*Ordinary Least Squares Regression* used in *Phenomenological Modeling*); 'deSolve' (ODE solver used in *Phenomenological Modeling*); and 'extRemes' (*Extreme Value Statistics*). The script and data used in the illustrative case study are

included as Supplementary Materials.

3. Illustrative Case Study

We illustrate the *NLTS* data diagnostics framework with a case study of saltwater intrusion into coastal wetlands in Everglades National Park, Florida (Figure 2). Saltwater intrusion is governed by interaction between tidal dynamics and the flow of freshwater through the Everglades [47]. Historically, freshwater discharge from Lake Okeechobee flowed southward through the Everglades as a slow-moving 48km-wide sheet of water first over sawgrass prairie in the northern portion, and then over and through peat land in the southern portion before draining into the mangrove keys of Florida Bay at Florida's southern tip [48]. In 1947, the federal Central and Southern Florida Flood Control Project authorized construction of massive water-control structures diverting Everglades water to urban development and re-shaping flows to drain Everglades wetlands converted to agricultural production. These structures blocked water flow through the Everglades and flattened the southern terrain as the 'ridge-and-slough' landscape—created by natural flows channeling through peat land—was leveled likely due to reduced sediment transport [48]. Experts hypothesize that slight increases in sea level along with large reductions in natural freshwater discharge and flattened terrain render the Everglades more susceptible to saltwater intrusion and ensuing environmental degradation [47, 48].

Stabenau *et al.* (2011) report that water levels in coastal Everglades marshes correlate with global climatic teleconnections—specifically the El Niño Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO) [47]. In this illustration, we empirically test for a wider range of hypothesized causal interactions among variables operating in Florida bay: Everglade water and salinity levels, sea level, ENSO, NAO, and an additional key teleconnection—the Atlantic Multidecadal Oscillation (AMO).

We obtained daily mean sea levels (*mm*) observed at the Key West tide gauge situated in Florida Bay [49] (Figure 2), daily median water levels (*feet*, NAVD88) [50] and salinity levels (*uS/cm*) [51] reported at the P_{36} station situated in the coastal wetlands at the southern tip of the Park, monthly Kaplan and Reynolds El Niño 3.4 index (ENSO) [52], the monthly Kaplan SST AMO index [53], and the monthly NAO index [54]. The sea level data, well level data, and well salinity data were aggregated to average monthly values for consistency with the climate indices. We converted the data to anomalies by subtracting the mean from each time series. The data set extends from January 1992 through December 2012 (252 observations).

Signal processing uncovered sufficiently strong signals in observed anomalies to justify moving forward with *NLTS*. Figure 3 plots the time series (solid lines), separated signals (dashed curves) and residual noise (dotted curves) and Table 1 reports signal strength as the percentage of variation explained in the corresponding anomaly, and the strength of constituent trends and oscillations. All signals (except for the NAO index) explain a larger percentage of variation than unstructured noise, ranging from a high of 77% (ENSO) to a low of 55% (sea level and well salinity). The signals for sea level, well level, and well salinity are dominated by annual oscillations accompanied by fainter lower-frequency oscillations. The annual oscillation in well level corresponds to the May-October wet season each year. The ENSO signal is dominated by lower-frequency (multi-annual) oscillations. The AMO signal is dominated by a nonlinear trend with fainter multi-annual oscillations. The NAO index provided a very weak signal perhaps because the time series was too short to resolve low-frequency oscillations.

We used (detrended) signals to successfully reconstruct low-dimensional attractors for sea level, well level, well salinity, ENSO, and AMO with time-delay embedding. The attractors demonstrate noticeable torus-type regularity characterized by non-repeating oscillations (Figure

4). The estimated embedding dimension was $M = 4$ for all signals (except for sea level where $M = 3$) so that the reconstructed attractors depicted in Figures 4b-e are three-dimensional projections. This offers preliminary empirical evidence of deterministic low-dimensional nonlinear system dynamics.

We first tested the null hypothesis that nonrepeating cycling characterizing the empirically-reconstructed attractors is generated by linear stochastic dynamics with the *AAFT* algorithm (Table 2) and then re-tested with the *PPS* algorithm (Table 3). We specified a significance level of 95% which generated 99 surrogates when using correlation dimension and Lyapunov exponent as discriminating statistics (two-sided hypothesis tests), and 49 surrogates when using predictive skill (one-sided test). Surrogate data testing uniformly rejects the null hypothesis for the environmental signals (sea level, well level and well salinity) for both algorithms. However, the results for the teleconnection signals (AMO and ENSO) are mixed. When using *AAFT* surrogates, the null hypothesis is accepted for AMO for all discriminating statistics, and rejected for ENSO only for predictive skill (Table 2). We proceeded with nonlinear diagnostics of the teleconnection signals because re-testing with *PPS* surrogates largely reversed these results (Table 3).

We applied *Convergent Cross Mapping* (*CCM*) to test the following hypotheses: (1) Sea level in Florida Bay drives well level and well salinity in the coastal Everglades; (2) AMO drives sea level, well level and well salinity; and (3) AMO and ENSO bicausally interact. These hypotheses are grounded in the literature: Salt water intrusion links ocean-surface salinity with salinity of inland water sources, and sea level with inland water level [47]. AMO and ENSO may share a teleconnective relationship through the atmosphere [55, 56], and both influence inland precipitation and thus well levels [57].

CCM provides empirical evidence in favor of the above hypotheses. Each of the four plots in Figure 5 uses the empirically-reconstructed attractor for one of the variables to test the extent to which that variable skillfully cross-predicts the others, and thus is a driver of them. Skillful cross-prediction is indicated when the correlation coefficient (vertical axes) converges to a high value as the number of points on the attractor used to cross predict increases (horizontal axes) [24].

Figure 5a shows that the empirically-reconstructed attractor for sea level skillfully cross-predicts well salinity and well level with a saturated CCM correlation coefficient of $\rho \approx 0.8$ and $\rho \approx 0.7$, respectively, providing evidence that sea level in Florida Bay drives well water and salinity. Well water drives well salinity through dilution as expected (5b). AMO is a weak driver of sea level (5c, black curve) since the correlation coefficient saturates at ~0.4, a slightly stronger driver of well water level and salinity level, and a strong driver of ENSO with $\rho \approx 0.8$. Finally, ENSO has a strong bi-causal interaction with AMO as evidenced by a saturated $\rho \approx 0.8$ (5d), but drives the environmental variables only indirectly through its interaction with AMO. These causal interactions are summarized in a network diagram where arrows indicate directions of causality, and color denotes strength (low to strong) of causality (Figure 5e).

Next, we succeeded in reproducing the empirically-reconstructed attractors by solving the following deterministic system of ordinary differential equations with first-order polynomial interaction terms:

$$\left.\begin{aligned}
\frac{dx}{dt} &= \alpha_0 + \alpha_1 w + \alpha_2 z + \alpha_3 zw + \alpha_4 y + \alpha_5 yw + \alpha_6 yz + \alpha_7 yzw + \alpha_8 x + \alpha_9 xw + \alpha_{10} xz + \alpha_{11} xzw + \alpha_{12} xy + \alpha_{13} xyw + \alpha_{14} xyz + \alpha_{15} xyzw \\
\frac{dy}{dt} &= \beta_0 + \beta_1 w + \beta_2 z + \beta_3 zw + \beta_4 y + \beta_5 yw + \beta_6 yz + \beta_7 yzw + \beta_8 x + \beta_9 xw + \beta_{10} xz + \beta_{11} xzw + \beta_{12} xy + \beta_{13} xyw + \beta_{14} xyz + \beta_{15} xyzw \\
\frac{dz}{dt} &= \gamma_0 + \gamma_1 w + \gamma_2 z + \gamma_3 zw + \gamma_4 y + \gamma_5 yw + \gamma_6 yz + \gamma_7 yzw + \gamma_8 x + \gamma_9 xw + \gamma_{10} xz + \gamma_{11} xzw + \gamma_{12} xy + \gamma_{13} xyw + \gamma_{14} xyz + \gamma_{15} xyzw \\
\frac{dw}{dt} &= \xi_0 + \xi_1 w + \xi_2 z + \xi_3 zw + \xi_4 y + \xi_5 yw + \xi_6 yz + \xi_7 yzw + \xi_8 x + \xi_9 xw + \xi_{10} xz + \xi_{11} xzw + \xi_{12} xy + \xi_{13} xyw + \xi_{14} xyz + \xi_{15} xyzw
\end{aligned}\right\} \quad (3)$$

where variables are sea level (x), well level (y), well salinity (z), and AMO (w); the left-hand side variables are the respective growth rates approximated with 4th order centered finite differences; and α_i, β_i, γ_i, and ξ_i (i = 0,…,15) are constant coefficients estimated with *Ordinary Least Squares* (*OLS*) regression.

NLTS diagnostics informed the specification of the phenomenological model as follows: First, the embedding dimensions estimated for three of the four signals (M = 4) indicated that the phenomenological model must be at least four dimensional to contain the system attractor. Second, *CCM* results provided empirical evidence that the included four variables interact in the same real-world dynamic system. Finally, empirically-reconstructed system attractors (Figure 4) provided a real-world benchmark of long-run system dynamics that a specification with first-order polynomial interactions succeeded in reproducing (as discussed below).

The phenomenological model provided a good fit to the finite differences calculated from the data (Table 4). Computed *Nash-Sutcliffe Efficiencies* (*NSE*) for sea level (0.78), well level (0.83), well salinity (0.88), and AMO (0.84) far exceed the 0.65 threshold generally held to constitute a reasonable fit [41]. Following Ritter and Muñoz-Carpena [41], block-bootstrapping

the observed and predicted values generated 95% confidence intervals of *NSE* for each equation and estimated *Cumulative Distribution Functions* (*CDFs*). These were used to strongly reject the null hypothesis that *NSE* falls below the 0.65 minimum model quality threshold with probability values $p = 0$ [41]. Table 4 also reports the *OLS* estimated coefficients with *t*-values and significance levels.

The solution of the phenomenological model for simulated values of the sea level, well level, well salinity, and AMO signals showed that the simulated signals share a dominant cycle of ~15 months—in the neighborhood of the 12 month cycle exhibited by the observed signals. We applied time-delay embedding to the simulated sea level signal to reconstruct the system dynamics of the phenomenological model. The simulated attractor (Figure 6) bears striking resemblance to the empirically-reconstructed attractor (Figure 4a), providing additional evidence that a deterministic, nonlinear and low-dimensional model corresponds to real-world dynamics in this case study.

Finally, we modeled probabilistically the unstructured noise separated from the observed time-series anomalies. Figure 7 shows return-level plots (log scale) for each variable estimating the months expected ('return times') before positive (left column) and negative (right column) noise levels exceed selected thresholds ('return levels') reported in Table 5. The return-level plots represent good fits of the Generalized Pareto Distribution to noise levels since each falls within bootstrapped 95% confidence intervals (dashed lines).

Consider, for example, noise separated from observed sea level anomalies. The positive noise level (for which the observed anomalies exceed the signal) exceeded the selected threshold of 25 *mm* a total of 77 times at a monthly rate of 30% (Table 5). The negative noise level (for which the observed anomalies rest below the signal) exceeded the threshold of 25 *mm* a total of

81 times at a monthly rate of 32%. The 6-month, year, and two-year return-levels for positive noise levels exceeding the threshold are calculated to be 58 *mm*, 88 *mm*, and 110 *mm*, respectively (Figure 6a).

4. Discussion

NLTS data diagnostics provide evidence of real-world dynamic behavior, and are intended to inform—not replace—mechanistic decision-support modeling needed to provide theory-based explanations for diagnosed behavior. Signal processing identifies structured variation in observed data whose cyclical patterns should be well approximated by a mechanistic model. *Phase Space Reconstruction* provides a geometric picture of real-world long-term system dynamics that should be reproduced by a mechanistic model. The estimated *embedding dimension* indicates the minimum system dimensionality required to reproduce diagnosed real-world dynamics in subsequent modeling. *Phenomenological modeling* tests whether diagnosed real-world dynamics can be reproduced by a deterministic, nonlinear, and low-dimensional system of ODEs composed of polynomial interactions of a determined order. If so, an ensuing decision-support model need not include more complex polynomial interactions to successfully simulate diagnosed real-world dynamics. Once a mechanistic model has been specified, the approach outlined by Baker et al. (1996) can be applied to estimate coefficients in a known dynamical system.

5. Conclusions

This paper is motivated by the increasing call for modelers to demonstrate that decision-support models used in environmental policy correspond to the reality facing policy-makers. We propose a two-pronged burden of proof: (1) modelers produce persuasive empirical evidence of real-world dynamic behavior; and (2) modelers demonstrate that their models skillfully simulate

this behavior.

Principles of randomness from the philosophy of science literature debunk a common presumption that irregular and random-appearing observed data provide evidence of inherently random (indeterministic) real-world biophysical processes requiring stochastic decision-support modeling. Nonlinear deterministic processes also can produce mathematically random output, and modelers cannot go backwards from observing random-appearing data to prove which type of real-world process generated them.

We propose a pre-modeling data diagnostics framework based on *Nonlinear Time Series* methods designed to detect whether deterministic real-world environmental dynamics are concealed in observed data (first prong of the burden of proof), and to guide specification and testing of decision-support models for correspondence with diagnosed real-world dynamics (second prong of the burden of proof).

A strong signal separated from observed data is tested for low-dimensional, nonlinear and deterministic dynamic structure. If this dynamic structure is identified, an empirically-reconstructed attractor passing surrogate-data testing provides a geometric picture portraying systematic real-world environmental dynamics that a well specified model should simulate in post-model specification auditing. The attractor can be used further for empirical causal network analysis to test for variables interacting in the same dynamic system, and for state-space forecasting. Phenomenological (data-driven) modeling of interacting variables sheds light on whether diagnosed real-world dynamics can be reproduced by a deterministic, nonlinear, and low-dimensional decision-support model. Noise separated from observed data can be modeled probabilistically with *Extreme Value Statistics* to account for extreme events occurring outside of detected temporal behavioral patterns. Return-level plots can be used to estimate the time

periods expected before a noise level exceeding a given threshold is realized.

We suggest that environmental modelers: (1) initially test for systematic dynamic behavior in observed data before resorting to conventional exploratory stochastic approaches that would fail to detect this valuable information; and (2) use non-linear dynamics diagnosed in the observed data as a rigorous benchmark for specification and testing of ensuing environmental decision-support models.

Acknowledgements. RH acknowledges support from NIFA, RMC acknowledges support from the UF Water Institute Faculty Fellowship.

Supplementary Materials. The data and R-script used in the case study can be downloaded from…

5. References

[1] A. Saltelli and S. Funtowitz, "When all models are wrong," *Computer Modeling,* vol. Winter 2014, pp. 79-85, 2014.

[2] N. Oreskes, K. Shrader-Frechette, and K. Belitz, "Verification, validation, and confirmation of numerical models in the earth sciences," *Science,* vol. 263, pp. 641-646, 1994.

[3] *Joint Research Centre, The European Commission, Sensitivity analysis, sensitivity auditing and impact assessment, EU website. Available: https://ec.europa.eu/jrc/en/research-topic/sensitivity-analysis-sensitivity-auditing-and-impact-assessment. Accessed 2015 Jan 10*.

[4] E. Rykiel, "Testing ecological models: the meaning of validation," *Ecological Modeling,* vol. 90, pp. 229-244, 1996.

[5] G. Hornberger and R. Spear, "An approach to the preliminary analysis of environmental systems," *Journal of Environmental Management,* vol. 12, pp. 7-18, 1981.

[6] L. Uusitalo, A. Lehikoinen, I. Helle, and K. Myrberg, "An overview of methods to evaluate uncertainty of deterministic models in decision support," *Env. Model. & Software,* vol. 63, pp. 24-31, 2015.

[7] G. Feder, "Pesticides, information, and pest management under uncertainty," *American Journal of Agricultural Economics,* vol. 61, pp. 97-103, 1979.

[8] B. Dixon and R. Howitt, "Resource production under uncertianty: A stochastic control approach to timber harvest scheduling," *American Journal of Agricultural Economics,* vol. 62, pp. 499-507,

1980.

[9] M. Johnson and E. Pasour, "An opportunity cost view of fixed asset theory and the overproduction trap," *American Journal of Agricultural Economics,* vol. 63, pp. 1-7, 1981.

[10] L. Karp, "Methods for selecting the optimal dynamic hedge when production is stochastic," *American Journal of Agricultural Economics,* vol. 69, pp. 647-657, 1987.

[11] B. Horan, "The statistical character of evolutionary theory," *Philosophy of Science,* vol. 61, pp. 76-95, 1994.

[12] R. Huffaker, "Deterministic modeling without unwarranted apology," *Review of Agricultural Economics,* vol. 20, pp. 502-512, 1998.

[13] P. Glendinning, *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations*. Cambridge: Cambridge University Press, 1994.

[14] A. Medio, *Chaotic Dynamics*: Cambridge University Press, 1993.

[15] T. Schreiber, "Interdiciplinary application of nonlinear time series methods," *Physics Reports,* vol. 308, pp. 1-64, 1999.

[16] L. Larsen, T. Eppinga, and T. Coulthard, "Exploratory modeling: Extracting causality from complexity," *EOS,* vol. 95, pp. 285-292, 2014.

[17] J. Elsner and A. Tsonsis, *Singular Spectrum Analysis*. New York: Plenum Press, 2010.

[18] M. Ghil, M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann*, et al.*, "Advanced spectral methods for climatic time series," *Reviews of Geophysics,* vol. 40, pp. 1-41, 2002.

[19] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure*. New York: Chapman & Hall/CRC, 2001.

[20] R. Vautard, "Patterns in time: SSA and MSSA in analysis of climate," in *Analysis of Climate Variability*, H. von Storch and A. Navarra, Eds., ed: Springer, 1999.

[21] H. Kantz and T. Schreiber, *Nonlinear Time Series Anaysis*. Cambridge: Cambridge University Press, 1997.

[22] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Farmer, "Testing for nonlinearity in time series: The method of surrogate data," *Physica D,* vol. 58, pp. 77-94, 1992.

[23] R. Katz, "Statistics of extremes in climate change," *Climatic Change,* vol. 100, pp. 71-76, 2010.

[24] G. Sugihara, R. May, Y. Hao, H. Chih-hao, E. Deyle, M. Fogarty*, et al.*, "Detecting causality in complex ecosystems," *Science,* vol. 338, pp. 496-500, 26 October 2012.

[25] G. Baker, J. Gollub, and J. Blackburn, "Inverting chaos: Extracting system parameters from experimental data," *Chaos,* vol. 6, pp. 528-533, 1996.

[26] M. Kot, Schaffer, W., Truty, G., Graser, D., Olsen, L. , "Changing criteria for imposing order," *Ecological Modeling,* vol. 43, pp. 75-110, 1988.

[27] G. Greco, R. Rosa, G. Beskin, S. Karpov, L. Romano, A. Guarnieri*, et al.* (2011, Evidence of Deterministic Components in the Apparent Randomness of GRBs: Clues of a Chaotic Dynamic. *Scientific Reports*.

[28] T. Brown, "Measuring Chaos Using the Lyapunov Exponent," in *Chaos Theory in the Social Sciences: Foundations and Applications*, E. Kiel and E. Elliott, Eds., ed Michigan: University of Michigan, 1996, pp. 53-66.

[29] S. Strogatz, *Nonlinear Dynamics and Chaos*. New York: Addison-Wesley Publishing Company, 1994.

[30] P. Glendinning, *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations*. Cambridge, UK: Cambridge University Press, 1994.

[31] D. Kaplan and L. Glass, *Understanding Nonlinear Dynamics*. New York: Springer, 1995.

[32] G. Williams, *Chaos Theory Tamed*. Washington D.C.: John Henry Press, 1997.

[33] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence* D. Rand, Young, L. , Ed., ed New York: Springer, 1980, pp. 366-381.
[34] J. Gleick, *Chaos*. Harrisonburg, Virginia: R.R. Donnelley & Sons Company, 1987.
[35] M. Small and C. Tse, "Applying the method of surrogate data to cyclic time series," *Physica D,* vol. 164, pp. 187-201, 2002.
[36] M. Small and C. Tse, "Detecting determinism in time series: The method of surrogate data," *IEEE Transactions on Circuits and Systems,* vol. 50, pp. 663-672, 2003.
[37] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D,* vol. 142, pp. 346-382, 2000.
[38] D. Kugiumtzis, "Surrogate Data Test on Time Series," in *Modelling and Forecasting Financial Data, Techniques of Nonlinear Dynamics*, A. Soofi and L. Cao, Eds., ed: Kluwer Academic Publishers, 2002.
[39] D. Kugiumtzis, "Test Your Surrogate Data before You Test for Nonlinearity," *Physical Review E,* vol. 60, pp. 2808 - 2816, 1999.
[40] J. Nash and J. Sutcliffe, "River flow forecasting through conceptual models, Part 1: A discussion of principles," *J. Hydrol.,* vol. 70, pp. 90255-6, 1970.
[41] A. Ritter and R. Munoz-Carpena, "Performance evaluation of hydrologic models: statistical significance for reducing subjectivity in goodness-of-fit assessments," *Journal Of Hydrology,* vol. 480, pp. 33-45, 2013.
[42] R. Harmel, P. Smith, K. Migliaccio, I. Chaubey, K. Douglas-Mankin, B. Benham*, et al.*, "Evaluating, interpreting, and communicating hydrologic/water quality model performance considering intended use: recommendations and review of best practices," *Env. Model. & Software,* vol. 57, pp. 40-51, 2014.
[43] L. Wei-Dong, K. Ren, S. Meunier-Guttin-Cluzel, and G. Gouesbet, "Global vector-field reconstruction of nonlinear dynamical systems from a time series with SVD method and validation with Lyapunov exponents," *Chinese Physics,* vol. 12, pp. 1366-1373, 2003.
[44] G. Gouesbet and J. Macquet, "Construction of phenomenological models from numerical scalar time series," *Physica D,* vol. 58, pp. 202-215, 1992.
[45] B. Mevik and R. Wehrens, "The pls package: Principal component and partial least squares regression in R," *Journal of Statistical Software,* vol. 18, pp. 1-24, 2007.
[46] E. Gilleland. (2015). *R package extRemes. Available: http://www.assessment.ucar.edu/toolkit/. Accessed 2016 Jan 2*.
[47] E. Stabenau, V. Engel, J. Sadle, and L. Pealstine, "Sea-level rise: Observations, impacts, and proative measures in Everglades National Park," *Park Science,* vol. 28, 2011.
[48] H. Fling, N. Aumen, T. Armentano, and F. Mazzotti, "The role of flow in the everglades landscape," The University of Florida, IFAS Extension, CIR 1452, 2004.
[49] *University of Hawaii Sea Level Center (UHSLC). Available: http://uhslc.soest.hawaii.edu/data/download/fd#uh242. Accessed 2015 Feb 28.*
[50] *Everglades Depth Estimation Network (EDEN). Available: http://www.sofia.usgs.gov/eden/eve/index.php?site_list[]=P35&water_le. Accessed 2015 Feb 28.*
[51] *DBHYDRO (South Florida Water Management District). Available: http://my.sfwmd.gov/portal/page/portal/xweb%20environmental%20monitoring/dbhydro%20application. Accessed 2015 Dec 27.*
[52] *KNMI Climate Explorer. Available: http://climexp.knmi.nl/selectindex.cgi?id=someone@somewhere. Accessed 2015 Dec 15.*
[53] *Earth System Research Laboratory. Available:*

http://www.esrl.noaa.gov/psd/data/timeseries/AMO/ *Accessed 2015 Dec 15.*

[54] *Climate Prediction Center (NOAA). Available:* http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/nao.shtml. *Accessed 2015 Dec 15.*

[55] J. Park and G. Dusek, "ENSO components of the Atlantic multidecadal oscillation and their relation to North Atlantic interannual coastal sea level anomalies," *Ocean Science,* vol. 9, pp. 535-543, 2013.

[56] D. Enfie!d and D. Mayer, "Tropical Atlantic sea surface temperature variability and its relation to El Nifio-Southern Oscillation," *Jounal of Geophysical Research,* vol. 102, pp. 929-945, 1997.

[57] D. Enfield, A. Mestas-Nunez, and P. Trimble, "The Atlantic multi-decadal oscillation and its relation to rainfall and river flows in the continental US," *Geophysical Research Letters,* vol. 28, pp. 2077-2080, 2001.

Table 1. Signal Processing: Percent Variance in Observed Data Explained (m = month)[a]

	Signal	Trend	60m	42m	40m	32m	24m	20m	18m	13m	12m
Sea level	55%[b]	--	--	--	--	--	--	--	--	--	55%
Well level	67%	--	--	--	--	--	10%	5%	--	3%	49%
Salinity	55%	10%	8%	--	--	--	--	--	--	--	37%
ENSO	77%	--	--	18%	--	27%	--	--	32%	--	--
AMO	70%	43%	--	--	18%	--	--	--	--	--	9%

[a] Observed time series converted to anomalies from mean, [b] percent of variation explained in time series from its mean.

Table 2. *AAFT* Surrogate Data Results using Rank-Order Statistics[a]

	Signal	Surrogate (low)	Surrogate (high)	H_0
Sea level				
Correlation Dimension	1.16	1.41	2.15	reject
Lyapunov Exponent	0.02	0.43	1.80	reject
Predictive Skill	1.00	0.42	0.91	reject
Well level				
Correlation Dimension	2.17	2.21	2.95	reject
Lyapunov Exponent	0.09	0.13	0.76	reject
Predictive Skill	0.97	0.87	0.95	reject
Well Salinity				
Correlation Dimension	1.40	1.53	2.35	reject
Lyapunov Exponent	0.08	0.14	0.85	reject
Predictive Skill	0.99	0.89	0.96	reject
AMO				
Correlation Dimension	3.18	2.55	3.53	accept
Lyapunov Exponent	0.36	0.11	0.69	accept
Predictive Skill	0.92	0.82	0.94	accept
ENSO				
Correlation Dimension	2.78	2.06	4.28	accept
Lyapunov Exponent	0.11	0.11	0.68	accept
Predictive Skill	0.97	0.85	0.94	reject

[a] *AAFT* surrogates are used to test the null hypothesis that nonrepeating cycling characterizing the empirically-reconstructed attractors is generated by linear stochastic dynamics. The significance level is set at 95% resulting in 99 surrogates for correlation dimension and Lyapunov exponent (two-sided tests), and 49 surrogates for predictive skill (one-sided test).

Table 3. *PPS* Surrogate Data Results using Rank-Order Statistics[a]

	Signal	Surrogate (low)	Surrogate (high)	H_0
Sea level				
Correlation Dimension	1.16	2.39	4.43	reject
Lyapunov Exponent	0.02	0.58	1.21	reject
Predictive Skill	1.00	0.48	0.82	reject
Well level				
Correlation Dimension	2.17	3.60	4.76	reject
Lyapunov Exponent	0.09	0.74	1.36	reject
Predictive Skill	0.97	0.26	0.67	reject
Well Salinity				
Correlation Dimension	1.40	2.50	4.25	reject
Lyapunov Exponent	0.08	0.64	1.12	reject
Predictive Skill	0.99	0.45	0.79	reject
AMO				
Correlation Dimension	3.18	2.99	4.03	accept
Lyapunov Exponent	0.36	0.72	1.44	reject
Predictive Skill	0.92	-0.0009	0.44	reject
ENSO				
Correlation Dimension	2.78	1.49	2.56	reject
Lyapunov Exponent	0.11	0.73	1.47	reject
Predictive Skill	0.97	-0.12	0.91	reject

[a] *PPS* surrogates are used to test the null hypothesis that nonrepeating cycling detected in the empirically-reconstructed attractors is most likely due to a randomly-shifting limit cycle characteristic of stochastic linear dynamics. The significance level is set at 95% resulting in 99 surrogates for correlation dimension and Lyapunov exponent (two-sided tests), and 49 surrogates for predictive skill (one-sided test).

Table 4. OLS Estimated Parameters of Monomial Terms in Phenomenological Model

Monomials[a]	***x*.ODE**	***y*.ODE**	***z*.ODE**	***w*.ODE**
constant	3.12329	0.02722	-3.55485	0.00307
w	93.45660***[b]	0.70154***	-156.51218***	0.07050***
z	0.27441***	-0.00106***	-0.13949***	0.00028***
zw	0.05105	-0.00107	0.23256	0.00025
y	-37.04003***	-0.12633***	47.62305***	-0.02121***
yw	58.56974	0.25390	-49.78074	0.09052
yz	0.03312	0.00075**	-0.10264	0.00009*
yzw	0.80485	0.00171	0.37027	-0.00052
x	0.17767***	0.00228***	-0.52836***	0.00034***
xw	-0.63018	0.00121	0.47792	0.00046
xz	-0.00023	0.00001**	-0.00009	0.00000
xzw	0.00796	0.00011***	-0.01640***	0.00000
xy	-0.01312	0.00092*	-0.03999	0.00000
xyw	1.93960	0.01926***	-3.16153***	0.00059
xyz	-0.00139	0.00002***	-0.00220***	0.00000**
Xyzw	0.00887	-0.00007	0.00566	-0.00001
NSE[c]	0.78	0.83	0.88	0.84
	[0.74-0.82]	[0.79-0.86]	[0.84-0.90]	[0.81-0.87]
p-value[d]	0.000***	0.000***	0.000***	0.000***

[a]Variables are sea level (*x*), well level (*y*), well salinity (*z*), and AMO (*w*). The notation '*.ODE' represents the fitted ordinary differential equation for the variable; [b] t values; [c] 95% confidence interval in brackets; [d] $p(NSE \leq 0.65)$; significance codes: ***lower than α=0.01, **lower than α=0.05, *lower than α=0.10.

Table 5. Modeling noise with Extreme Value Statistics

	Positive Noise	Negative Noise
Sea level		
threshold	25 (*mm*)	25 (*mm*)
exceedances	77	81
monthly rate	0.3	0.32
Well level		
threshold	0.15 (*feet, NAD88*)	0.15 (*feet, NAD88*)
exceedances	81	65
monthly rate	0.32	0.26
Well Salinity		
threshold	40 (*uS/cm*)	45 (*uS/cm*)
exceedances	71	85
monthly rate	0.28	0.33
AMO		
threshold	0.05	0.05
exceedances	72	80
monthly rate	0.29	0.32

Figure Captions

Figure 1. Framework for NLTS pre-modeling data diagnostics and data-driven modeling. We first separate an observed time series into signal (structured variation) and noise (unstructured variation) components, and use the separated signal to reconstruct the dynamics of the real-world system generating it. We test whether empirically-detected dynamics are deterministic, low-dimensional, and nonlinear, and then whether attractors reconstructed from multiple observed signals are causally interrelated. We apply this information to simulate empirically-detected dynamics with a phenomenological (data-driven) model composed of polynomial ordinary differential equations. We complete the diagnostics by modeling unstructured noise separated from the observed time series probabilistically with Return-level Plots. These data diagnostics inform the structure of mechanistic models capable of simulating real-world system dynamics.

Figure 2. Case study lay out showing the locations of the Okechobee Lake, Everglades and monitoring sites for data (Well P36 for water table elevation and salinity and Key West Tide Gauge for sea level elevation).

Figure 3. Plots of observed time series with separated signals and noise. Singular Spectrum Analysis separated each observed time series (black curves) into signal (dashed curves) and residual noise (dotted curves). All signals (except for the NAO index) explain a larger percentage of variation in their respective time series than unstructured noise, ranging from a high of 77% (ENSO) to a low of 55% (sea level and well salinity). The signals detected in sea level, well level, and well salinity are dominated by annual oscillations, the ENSO signal by lower-frequency (multi-annual) oscillations, and the AMO signal a nonlinear trend with fainter multi-annual oscillations. The NAO index has a very weak signal dominated by unstructured

noise.

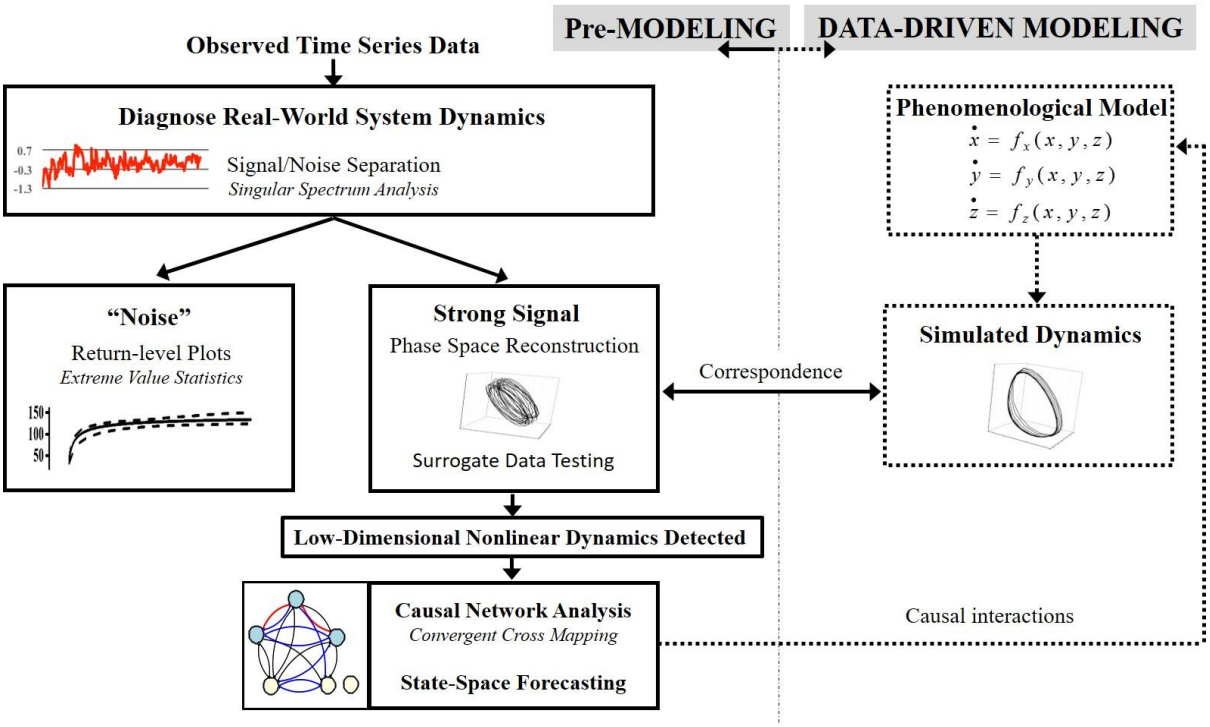
Figure 4. Attractors reconstructed from signals separated from observed time series data (trends removed). Attractors reconstructed from sea level (a), well level (b), salinity level (c), AMO (d) and ENSO (e) signals display 'noticeable regularity' characterized by torus-type non-repeated oscillations. This offers preliminary empirical evidence of deterministic low-dimensional nonlinear system dynamics.

Figure 5. Causal network analysis. We use *Convergent Cross Mapping* to ascertain whether attractors reconstructed from observed time series represent the dynamics of the same real-world system, and thus causally interact. The plots (a-d) show whether attractors reconstructed from sea level (a), well level (b), AMO (c), and ENSO (d) signals skillfully cross predict the other variables. An attractor skillfully cross predicts another variable when the Pearson correlation coefficient (vertical axes) converges to a high value as the number of points on the attractor used to cross predict increases (horizontal axes). In Figure 5a, for example, the attractor reconstructed from sea level skillfully cross-predicts well salinity as evinced by a saturated correlation coefficient of ~0.8 (blue dashed curve). The cross-predictions in 4a-d are summarized in network diagram (4e) where arrows indicate directions of causality, and color denotes strength of causality. Red arrows denote a high-level saturated correlation coefficient interval (0.8, 1), blue arrows an intermediate interval (0.6, 0.79), and black arrows the lowest-level interval that we considered noteworthy (0.4, 0.59).

Figure 6. Simulated system attractor reconstructed from phenomenological model. The phenomenological ODE model is solved for simulated values of the sea level, well level, well salinity, and AMO signals. System dynamics are reconstructed by applying time-delay

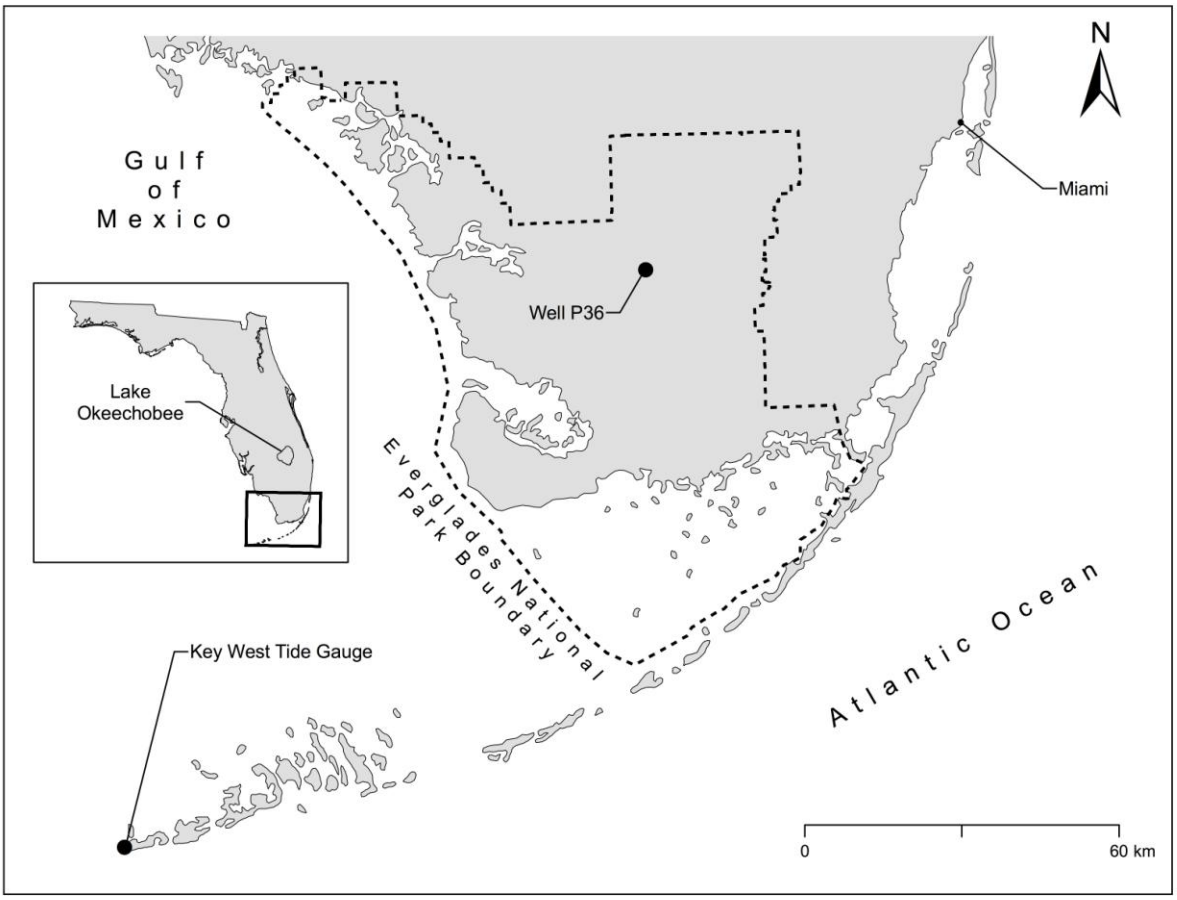
embedding to simulated sea levels. The model attractor bears striking resemblance to the empirically-reconstructed attractor (Figure 4a), providing evidence that a deterministic, nonlinear, and low dimensional ODE model was capable of corresponding to real-world dynamics.

Figure 7. Return-level plots for unstructured noise. The noise separated from the observed time-series anomalies is modeled probabilistically with Extreme Value Statistics. Positive (negative) noise is the amount by which the observed time series exceeds (falls below) the signal. The 'return-level' plots (log scale) estimate the months expected ('return times') before positive (left column) and negative (right column) noise levels exceed selected thresholds ('return levels'). Bootstrapped 95% confidence intervals for each variable are presented in dashed lines.



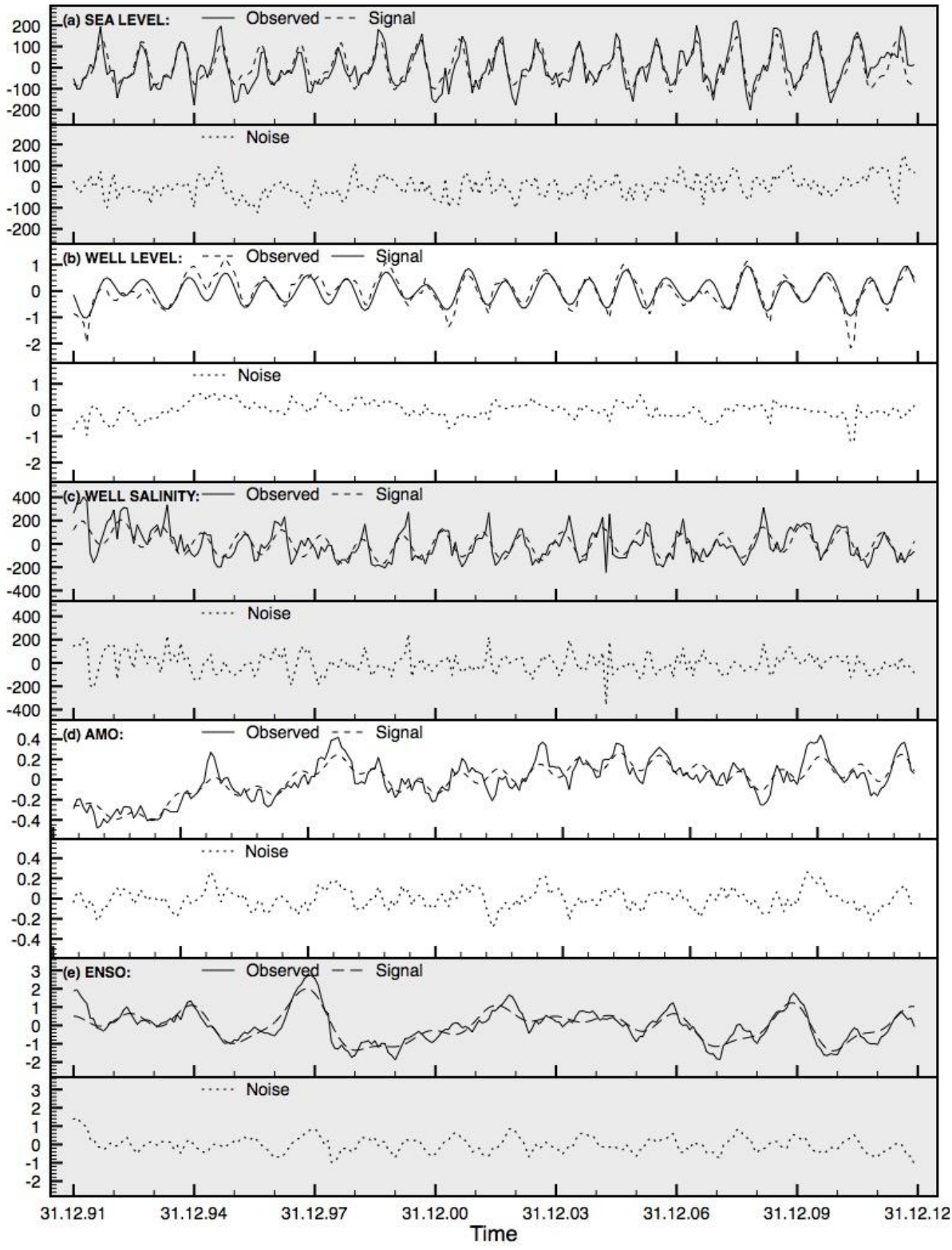
714

715 Figure. 1



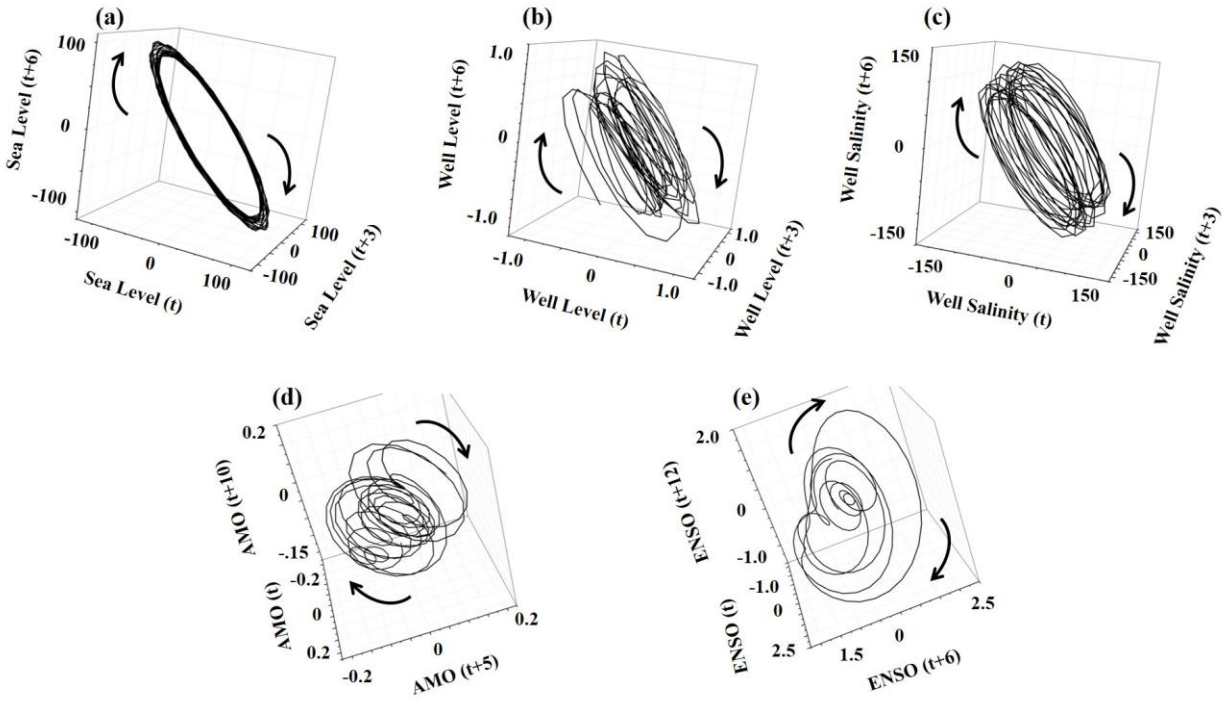
716

717 Figure 2.



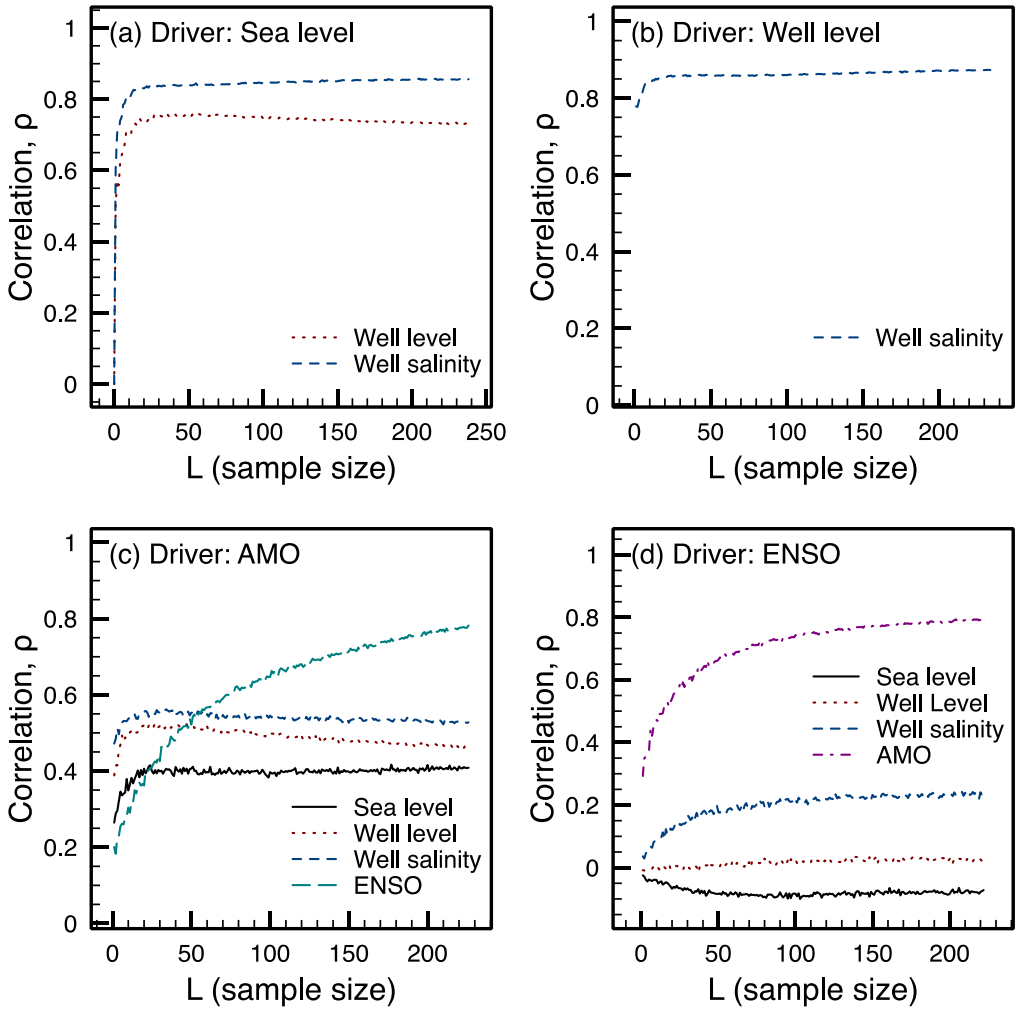
718

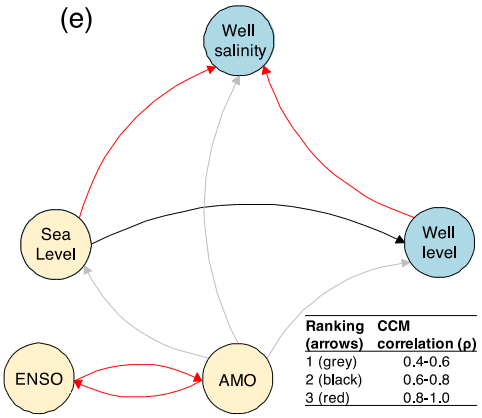
719 Figure. 3



720

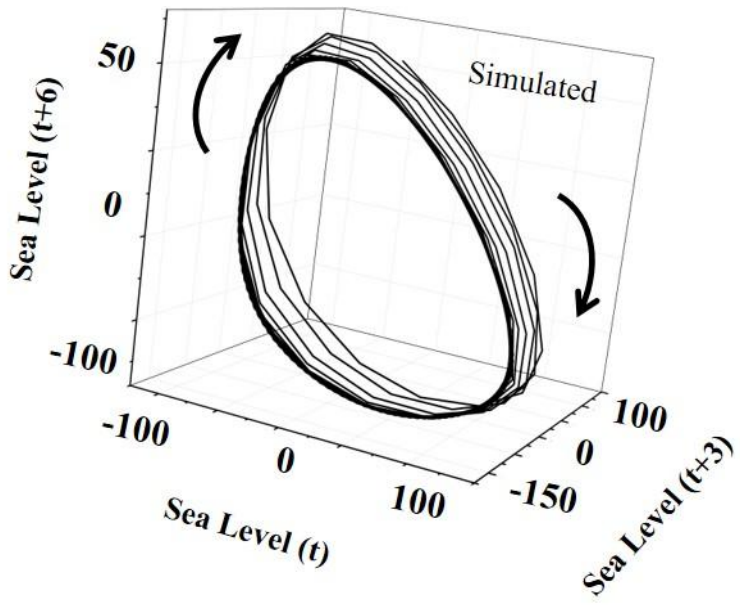
721 Figure 4.





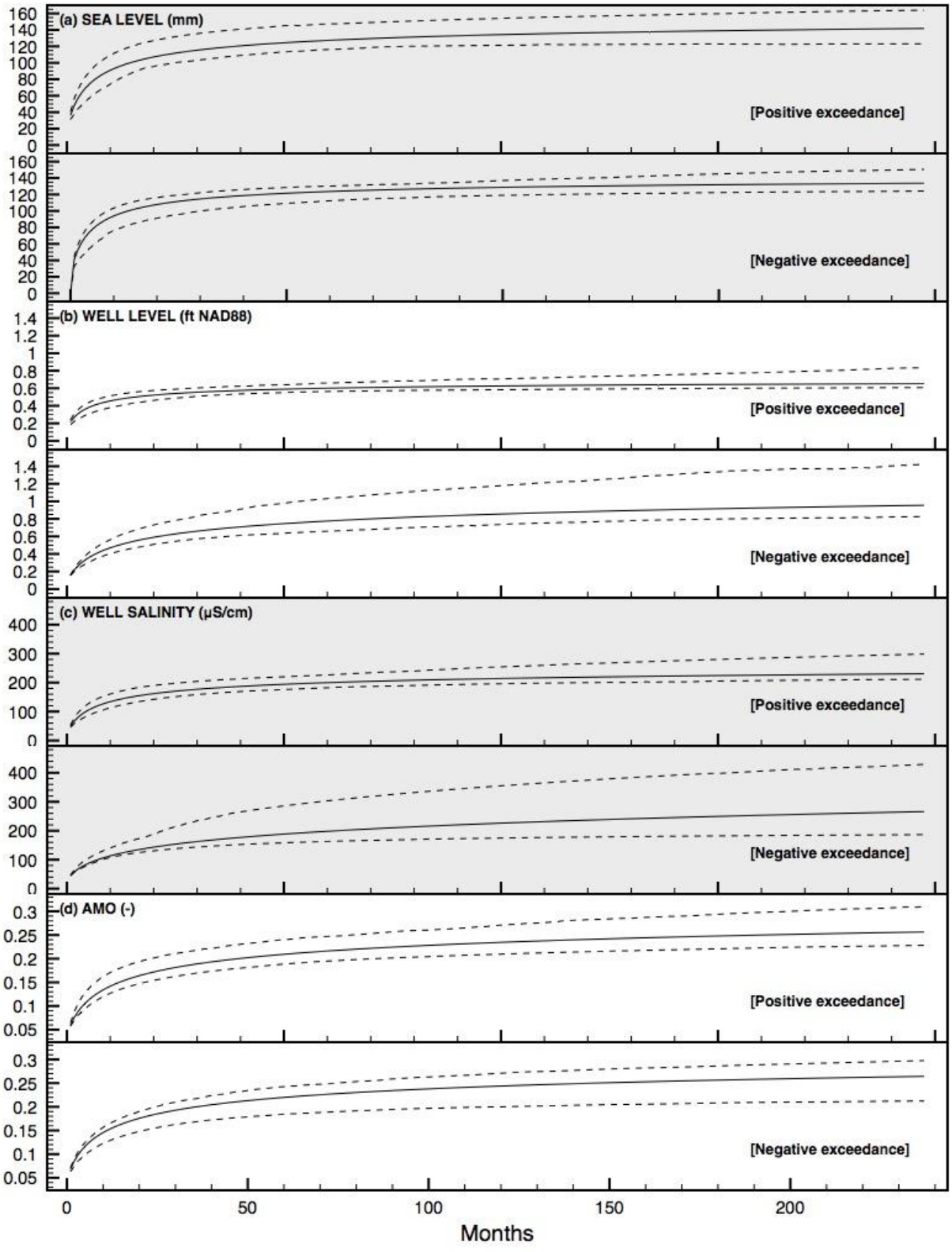
722

723 Figure. 5



724

725 Figure 6



726

727 Figure 7